Journal of Mechanics in Medicine and Biology
Vol. 13, No. 3 (2013) 1350026 (13 pages)

DOI: 10.1142/S0219519413500267

World Scientific
www.worldscientific.com

IDENTIFICATION OF KNEE FRONTAL PLANE KINEMATIC PATTERNS IN NORMAL GAIT BY PRINCIPAL COMPONENT ANALYSIS

NEILA MEZGHANI* and ALEXANDRE FUENTES
Laboratoire de recherche en Imagerie et Orthopédie
Centre de recherche du CHUM and
École de technologie supérieure, Montréal (Qc), Canada
**neila.mezghani@etsmtl.ca*

NATHALY GAUDREAULT
École de réadaptation, Sherbrooke University
Sherbrooke (Qc), Canada

AMAR MITICHE
Institut National de la Recherche Scientifique, INRS-EMT
Montréal (Qc), Canada

RACHID AISSAOUI, NICOLA HAGMEISTER
and JACQUES A. DE GUISE
Laboratoire de recherche en Imagerie et Orthopédie
Centre de recherche du CHUM and
École de technologie supérieure, Montréal (Qc), Canada

Received 29 June 2012
Revised 20 September 2012
Accepted 5 October 2012
Published 3 December 2012

The purpose of this study was to identify meaningful gait patterns in knee frontal plane kinematics from a large population of asymptomatic individuals. The proposed method used principal component analysis (PCA). It first reduced the data dimensionality, without loss of relevant information, by projecting the original kinematic data onto a subspace of significant principal components (PCs). This was followed by a discriminant model to separate the individuals' gait into homogeneous groups. Four descriptive gait patterns were identified and validated by clustering silhouette width and statistical hypothesis testing. The first pattern was close to neutral during the stance phase and in adduction during the swing phase (Cluster 1). The second pattern was in abduction during the stance phase and tends into adduction during the swing phase (Cluster 2). The third pattern was close to neutral during the stance phase and in abduction during the swing phase (Cluster 3) and the fourth was in abduction during both the stance and the swing phase (Cluster 4).

*Corresponding author.

Keywords: Gait pattern identification; frontal plane kinematics; clustering; principal component analysis.

1. Introduction

Gait analysis from biomechanical data, such as kinematic, kinetic, and electromyographic measurements, informs about an individual's locomotion function. A characterization of asymptomatic gait, also called normal gait, by a few patterns representative of distinct gait categories, is necessary to understand locomotion function.[1] However, such a characterization is difficult for two reasons. First, the biomechanical data are given in the form of a vector of measurements of high dimension for each subject, causing its analysis to suffer from the drawback of dimensionality.[2] Second, there is significant variability in the biomechanical data of asymptomatic gait.

Some studies have attributed kinematic data variability to differences in methodology (motion capture systems, walking speed, data processing, etc.),[3] but others have hypothesized it to be due to the presence of several different patterns in the asymptomatic gait data.[4] The multiplicity of knee kinematic patterns was backed by experiments with intracortical pins.[5] The high variability of gait data and the curse of dimensionality have forced most studies to use simple descriptions of the asymptomatic class, e.g., the mean of the available gait data or the data vector local extrema. Along this vein, there have been various biomechanical patterns to describe the normal three-dimensional (3D) kinematic patterns of the knee (flexion/extension, abduction/adduction, and tibial internal/external rotation) during functional activities such as level walking.[5–12] In most studies, for each anatomical plane, a single normal pattern with bands of deviation is shown as a reference to explain abnormalities in a patient's pattern.[13]

The problem we are facing when trying to determine characteristic patterns in curve bundles is basically one of data clustering because the purpose is to discover markers that can support clinical and treatment decisions within a large set of unlabeled patterns. As a result, pattern recognition and machine learning techniques are required to characterize gait patterns. There has been some work in the field of biomechanics in this respect. Various clustering techniques have been applied, including hierarchical clustering,[14–16] c-means,[17–20] and fuzzy clustering.[21] Hierarchical clustering was previously applied to distinguish gait patterns of young able-bodied subjects from those of able-bodied elderly men based on phasic and temporal parameters as well as 3D peak muscle powers.[14] It was also shown to be an effective method in identifying five distinct gait sub-patterns in young able-bodied men using peak 3D lower limb muscles powers as input.[15] Hierarchical c-means clustering algorithms were employed to derive gait patterns using temporal–distance parameters (stride length and step frequency/cadence).[17] Common clustering was used to classify spatiotemporal and kinematic data of patients following a stroke.[18] A c-means clustering method[19,20] and hierarchical clustering[16] have been

applied on kinematic data to identify gait pattern deviations in children with cerebral palsy. In this same pathologic population, fuzzy clustering was applied to group normal and pathological subjects using temporal−distance parameters.[21]

Previous studies have generally focused on pathological rather than asymptomatic gait because of the small sample size of the available asymptomatic dataset.[7] Moreover, they often used simple descriptors such as the mean of the measurement data or local information to identify patterns in the data. Sample data description can seriously limit the interpretation of the data.

For more meaningful gait data interpretation, we investigated the identification of asymptomatic patterns using the entire curve of the abduction/adduction (abd/add) angle versus the gait cycle (GC) percentage, rather than specific local information on these curves. The proposed method first reduced data dimensionality by projecting the original curve onto a subspace of significant components via principal component analysis (PCA). This was followed by a discriminant model to separate the individuals' gait into homogeneous groups. Our basic hypothesis was that distinct patterns can be found in the asymptomatic knee frontal plane kinematic gait data.

2. Methods

2.1. *Data acquisition*

One hundred and eleven (111) asymptomatic volunteers participated in this gait study. Institutional ethics approval was obtained and all participants gave their informed consent. The participants' descriptive characteristics are shown in Table 1.

Three-dimensional knee kinematic data were recorded while the participant was walking on a conventional treadmill at a self-selected comfortable speed. Kinematics data have been shown to be similar for treadmill and overground gaits.[22] Data collection was performed on each knee separately. For nine of the participants, measurements were only collected on one knee, giving a total of 213 dataset elements as shown in Fig. 2. Each participant took part in a 10-min treadmill walking adaptation period to ensure reproducible knee kinematics prior to data acquisition.

Table 1. Descriptive characteristics of subjects.

Parameter	Value
Number of subjects	111
Age (year)	36.3 ± 14.5
Weight (kg)	71.38 ± 13.6
Height (cm)	169.58 ± 8.6
Walking speed (m/s)	1.03 ± 0.18
Percentage of male/female	48%/52%

Fig. 1. Three-dimensional (3D) knee kinematic data acquisition system.

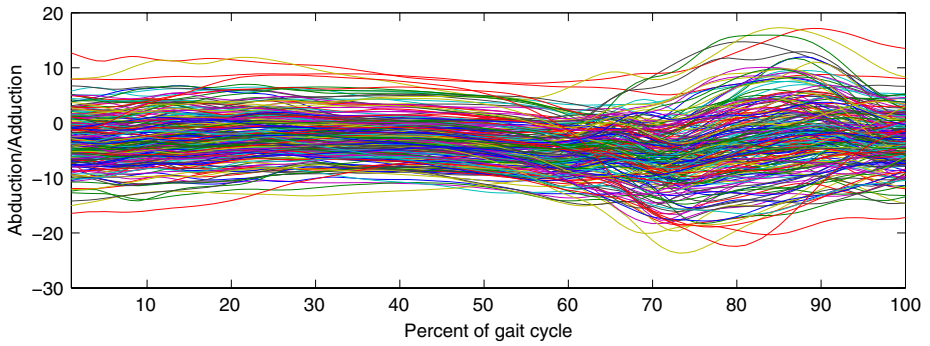
A knee marker attachment system, the KneeKG system[23,24] (EMOVI, Inc., Laval, Canada), was installed on the participant's knee to record the 3D kinematics as shown in Fig. 1. The position and orientation of the marker system, attached non-invasively onto an exoskeleton, were recorded using an electromagnetic motion tracking system (Fastrack, Polhemus, USA) at a sampling frequency of 60 Hz.

A number of representative GCs, generally 15, were averaged to obtain a mean pattern per subject. This was followed by interpolation and resampling from 1% to 100% of the GC, therefore giving 100 measurement points for each participant. Both the variability and high dimensionality are illustrated in Fig. 2, which shows the graph of a sample of 213 distinct asymptomatic abd/add curves, each composed of 100 measurement points.

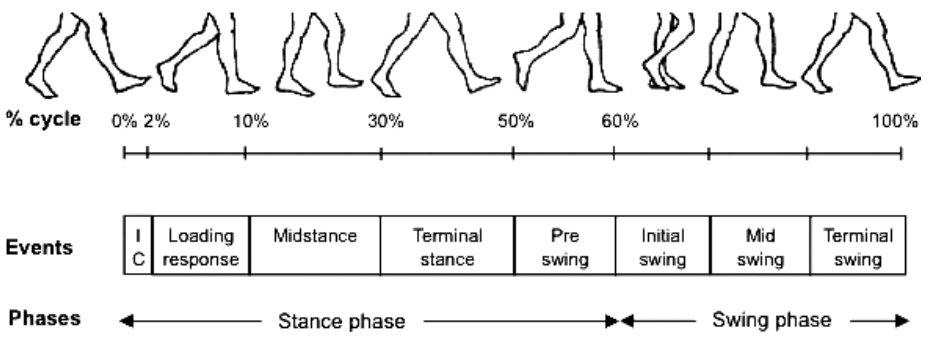
2.2. *Identification of gait patterns*

The purpose of this study was to extract meaningful patterns to separate the abd/add data into homogeneous groups. The data we worked with were all from the asymptomatic class; there was no *a priori* information such as manual labeling or other such ground truth that would allow distinguishing specific subclasses beforehand. In particular, we had no formal statistical model to represent the data, but rather, we worked with unlabeled data. The problem was therefore one of data clustering in an unsupervised framework.

In its general meaning, clustering is a non-parametric technique that consists of dividing a data sample into groups of similar elements. The elements in a group are more similar to each other than they are to the elements of the other groups. Clustering, as any other pattern classification technique, suffers from the problem of



(a)



(b)

Fig. 2. (a) Abduction/adduction (abd/add) angle curves over gait cycles (GCs) of the 213 asymptomatic subjects. The signals were interpolated and resampled from 1% to 100% (100 points) of the GC. (b) The GC phases.

dimensionality; unless very large, asymptotically infinite, the data is sparse in a high-dimensional space and therefore, difficult to separate into meaningful groups.[25] We addressed this problem by PCA — the original data was projected onto a low-dimensional subspace of relevant variables, called principal components (PCs), without affecting the inherent information contained about the meaningful groups in the sample. Details about the PCA method are given next.

2.3. *Principal component analysis*

PCA is an efficient technique for dimensionality reduction. It has been used in various fields of study,[26,27] including biomechanics.[28,29]

Given a set of data of p dimensions, PCA determines a linear subspace of dimension $k \leq p$ such that the projection of the data points on this subspace contains most of the variability of the data, i.e., most of the relevant information in the data. More precisely, let $\mathbf{X}_{(n\times p)} = \{x_{i,j} : 1 \leq i \leq n,\ 1 \leq j \leq p\}$ be the observed data

presented in matrix form. In this study, each individual contributes two rows in this matrix $\mathbf{X}$, one corresponding to the left knee angle and the other to the right knee angle. Therefore, n is even and $n/2$ is the number of individuals. Each column of the matrix corresponds to a GC percentage. There were 100 columns ($p = 100$) in our study; these correspond to the 100 points of the angle versus cycle percentage curve. The observation matrix $\mathbf{X}$ is generally expressed as a mean normalized to 0 and unit variance data. PCA produces linear combinations of the observations $\mathbf{X}_{(n\times p)}$ to determine the PC scores $\mathbf{Z}_{(n\times p)}$:

$$\mathbf{Z}_{(n\times p)} = \mathbf{X}_{(n\times p)} \times \mathbf{A}_{(p\times p)}. \tag{1}$$

According to Eq. (1), matrix $\mathbf{Z}$ contains the new variables (PCs) and is the product of the original data matrix and a linear transformation weight matrix $\mathbf{A}$. The columns of $\mathbf{A}$ are eigenvectors of the covariance matrix $\mathbf{\Sigma}$ under a spectral decomposition:[25]

$$\mathbf{\Sigma} = \mathrm{U}\mathbf{\Lambda}\mathrm{U}^T, \tag{2}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ is the diagonal matrix of the ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ and $\mathbf{U}$ is a $p \times p$ orthogonal matrix containing the eigenvectors.[25] Matrix $\mathbf{U}$ determines an orthonormal basis for the PCs.

The total variance is equal to the sum of the eigenvalues of the covariance matrix:[30]

$$\sum_{i=1}^{p} \mathrm{var}(PC_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \mathrm{trace}(\Sigma). \tag{3}$$

Therefore, the fraction

$$\sum_{i=1}^{k} \frac{\lambda_i}{\mathrm{trace}(\Sigma)} \tag{4}$$

gives the cumulative proportion of the variance explained by the first k PCs.

A powerful property of PCA is that the majority of the data variation can often be explained by just the first few PCs. As a result, the components corresponding to the smaller eigenvalues are dropped, affording a significant reduction in dimensionality without the loss of relevant information. The number of PCs needed to adequately describe a dataset can be determined using the portion of explained variance criteria.[31] In this study, a 90% trace criteria was used to determine the number of PCs to retain.

Dimensionality reduction was followed by a discriminant model, based on the PC's sign, to separate the individuals' gait into homogeneous groups. This model provides an effective and computationally efficient method to determine distinct patterns descriptive of the asymptomatic gait. Moreover, it is consistent with the assumption of Gaussian-distributed data, which served as the derivation of the PCA, namely to determine an orthonormal basis for the PCs.[25]

According to Eq. (1), the columns of $\mathbf{A}$, which are the eigenvectors of the covariance matrix of $\mathbf{X}$, are called PC coefficients or loadings. The component loadings are the correlation coefficients between the variables and the PCs. In this study, the variables were the GC percentage. As a general rule, variables with large loadings indicate that they are representative of the component, while small loadings suggest that they are not.[32] We used a parallel analysis to objectively determine a threshold estimate for significant loading coefficients. Parallel analysis was performed by running a simulation on a random variable. The total number of variable loadings (number of variables × number of extracted components) was multiplied by the selected significance level (e.g., $0.05 = 95$th percentile) providing an empirical estimate of the percentile loading.[33–35] As a result, we obtained an absolute value, the threshold, for the significant loading.

2.4. *Validation of gait patterns*

The purpose of gait pattern validation is to determine whether the dataset division resulted in well-distinguished groups. We verified the clustering quality based on the analysis of the silhouette width and with statistical evaluation by hypothesis testing.

The silhouette width is a confidence indicator of the membership of a sample to a cluster.[36] For a given cluster, C_j, $j = 1, 2, \ldots, c$, the silhouette width $s(i)$ of a sample $\mathbf{x}_i$ is:

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{5}$$

where a_i is the average distance between $\mathbf{x}_i$ and the elements in the cluster C_j, and b_i is the minimum of the average distances between the sample $\mathbf{x}_i$ and all of the samples clustered in C_k ($k = 1, 2, \ldots, c$ and $k \neq j$).

An analysis of variance was performed to verify group homogeneity. A post-hoc Tukey test was used to examine the differences between pairs of groups. The statistical analysis were conducted using SPSS 18.0 (Statistical Package for Social Sciences).[37] A p-value of 0.05 was set as the criterion for statistical significance.

3. Results

We used the portion of explained variance criteria to determine the proper number of PCs to retain. The first two PCs were retained as they explained 94% of the total variance. The first PC (PC1) accounted for 79% and the second PC (PC2) for 15%.

The patterns, representative of asymptomatic gait which the discriminant model identified, are the four mean patterns shown in Fig. 3. Figure 4 shows the corresponding clusters in matching colors.

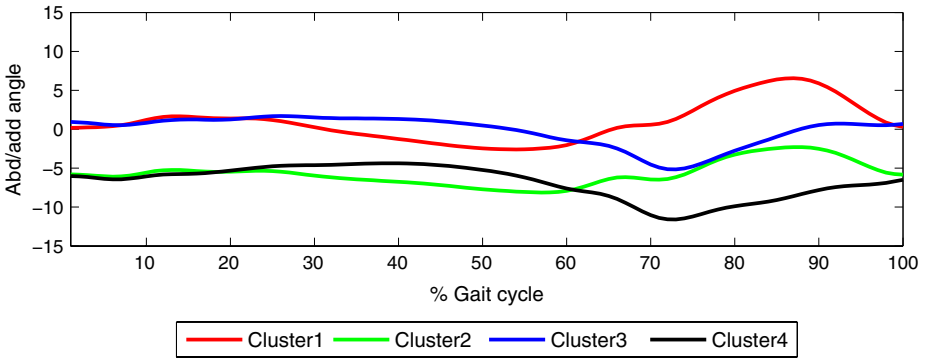


Fig. 3. The patterns representative of asymptomatic gait.

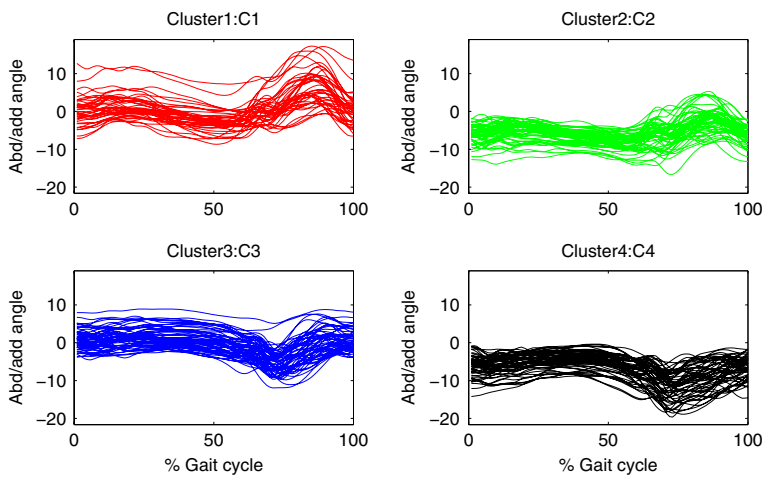
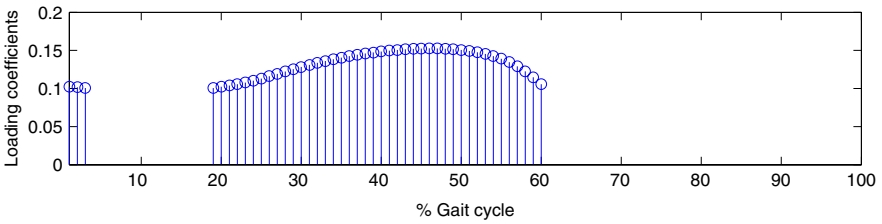


Fig. 4. The clustered abd/add samples corresponding to Cluster 1, Cluster 2, Cluster 3, and Cluster 4. Each curve corresponds to a subject in the original data space of Fig. 2.

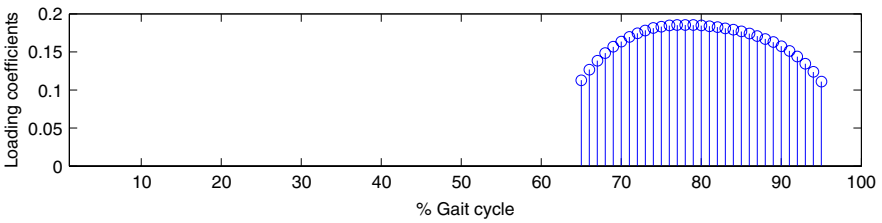
By looking at the loading coefficients in Fig. 5, we can see that the first PC represents the stance phase and the second PC represents the swing phase.

The silhouette plot which supported the validation is shown in Fig. 6. The plot accounts for each of the 202 observations. The observations are grouped per cluster, starting with Cluster 1 at the top; the silhouette values as horizontal bars are ordered to fit in each cluster. The silhouette width is positive for all the observations except for six samples (one in Cluster 1 and five in Cluster 2), which represent only about 2% of the dataset, i.e., 2% of the gait data sample have been assigned to a wrong cluster.

Table 2 summarizes the results of the statistical analysis. The results showed that the number of samples and the gender distribution within the four clusters are equivalent. There is no statistically significant difference between the four clusters



(a)



(b)

Fig. 5. Loading coefficients for the first (a) and second PCs (b). Only the highest loading >0.1 were included in the plot to make variable (GC percentage) interpretation easier. The first PC has higher loading factors during the stance phase of the GC (between 1% and 60%) and the highest loading factors of the second PC are located during the swing phase of the GC (between 60% and 100%).

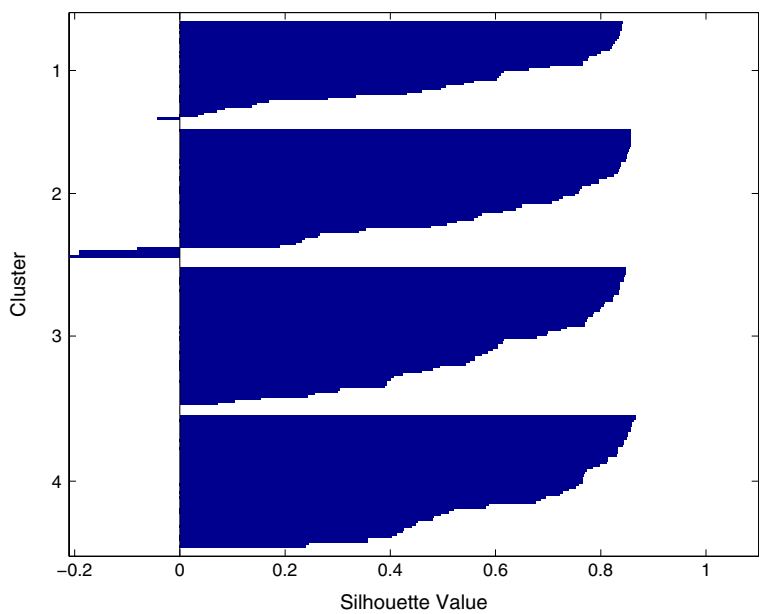


Fig. 6. Silhouette width bar graph of the clustered data. For each dataset element $\mathbf{x}_i, i = 1, 2, \ldots, 202$, a horizontal bar represent the silhouette width $s(i)$ to the cluster $C_j, j = 1, 2, 3, 4$. The width of a particular bar is proportional to the compactness. A negative silhouette value means that the observation is assigned to a wrong cluster.

Table 2. Demographic characteristics of subjects of the entire group and of the determined clusters.

	All samples of the data set	Cluster 1: C1	Cluster 2: C2	Cluster 3: C3	Cluster 4: C4
Number of samples	202	40	52	56	54
Age (year)	35.9 ± 14.2	39.1 ± 17.8	34.7 ± 13.0	39.2 ± 15.0*	31.6 ± 10.1*
Weight (kg)	71.7 ± 13.5	74.8 ± 10.4**	74.1 ± 16.1**	73.6 ± 13.8**	65.1 ± 10.5**
Height (cm)	169.7 ± 8.5	171.2 ± 8.4	168.5 ± 8.9	170.2 ± 8.9	169.3 ± 7.9
Walking speed (m/s)	1.1 ± 0.1	0.99 ± 0.1	1.1 ± 0.2	1.0 ± 0.2	1.1 ± 0.1
Percentage of male/ female	48%/52%	56%/44%	53%/47%	42%/58%	43%

Note: *$p < 0.05$, between C3 and C4.
**$p < 0.01$, between C1 and C4, C2 and C4, and C3 and C4.

with respect to height and walking speed. Clusters 1, 2, and 3 are similar in terms of anthropometric characteristics. However, the participants belonging to Cluster 4 are younger and have a lower mean body weight than the other three clusters.

4. Discussion and Conclusion

The silhouette graphical representation of Fig. 6 shows a good separation quality of the four clusters of the abd/add asymptomatic gait samples (Fig. 4). We noted that these clusters have a comparable number of samples (21% for Cluster 1, 28% for Cluster 2, 28% for Cluster 1, and 27% for Cluster 4). This is a desirable result because there is *a priori* no justification, statistical or clinical, for uneven repartition of the data points among the clusters.

By representing each cluster bundle by its mean curve (Fig. 3), we can also see the good separation between the four clusters. The four clusters demonstrated the occurrence of four specific patterns during both the stance phase (neutral versus abduction) and the swing phase (abduction versus adduction). The first pattern is close to neutral during the stance phase and in adduction during the swing phase (Cluster 1). The second pattern is in abduction during the stance phase and tends into adduction during the swing phase (Cluster 2). The third pattern is close to neutral during the stance phase and in abduction during the swing phase (Cluster 3) and the fourth pattern is in abduction during both the stance and the swing phases (Cluster 4). A representation by the total mean error band will inevitably mask these four clusters.

The loads of PC1 are high only for the variables corresponding to the GC stance phase as shown in Fig. 5. However, the PC2 loads are high only for the swing phase, also shown in Fig. 5. This affords an additional strong justification for the four clusters. More specifically, this means that most of the variability can be explained by events of clinical importance during the GC. Looking at PC1, the loading coefficients indicated that three specific events of the stance phase differentiate the patterns: (1) when the foot hits the ground (initial contact between 1% and 2% of

the GC, (2) during single limb stance, from mid-stance to the terminal stance, between 20% and 50% of the GC, and (3) during pre-swing, from 51% and 60%, when the heel is lifted off the ground (push off).

The identified clusters are in agreement with those of previous studies in the literature. Indeed, a knee adduction swing strategy (Clusters 1 and 2) has previously been reported[5,8,9,11,12] and a knee abduction swing strategy (Clusters 3 and 4) has also been observed.[5–7] The four clusters are consistent with the biomechanical experiments of Lafortune *et al.*,[5] who used intra-cortical pins to show that three out of five participants walked with close to neutral abd/add during the stance phase similar to Clusters 1 and 3, while the other two out of five walked with slight abduction similar to Clusters 2 and 4.

From a demographic point of view, the statistical test of Table 2 showed that there is no significant difference in walking speeds between the four clusters. They also showed the homogeneity of the anthropometric characteristics for Clusters 1, 2, and 3. These results emphasized that for these groups, the difference in waveform pattern cannot be attributed to age, weight, or height. The question regarding whether the lower weight observed for Cluster 4 could influence the waveform pattern is difficult to answer. The lower weight could be attributed to the fact that this group was also younger than the other three. Further experiments simultaneously analyzing the knee and the ankle using other parameters such as moments and powers would be needed to explain this difference.

In summary, this study has identified four distinct meaningful patterns in normal gait by PCA of knee abd/add angle data and a PC clustering model. The study can be extended to the analysis of the sagittal and transverse plane kinematics data.

Acknowledgments

The authors would like to thank the Programme MITACS of Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT), the Natural Sciences and Engineering Research Council (NSERC), and the Canada Research Chair in 3D Imaging and Biomedical Engineering. They would also like to thank Gerald Parent and Mathieu Charbonneau for their technical support.

References

1. Whittle MW, *An Introduction to Gait Analysis*, Butterworth-Heinemann, 2007.
2. Duda RO, Hart PE, Stork DG, *Pattern Classification*, Wiley-Interscience Publication, 2000.
3. Chau T, Young S, Redekop S, Managing variability in the summary and comparison of gait data, *J Neuroeng Rehabil* **22**(2):22–42, 2005.
4. Winter DA, Kinematic and kinetic patterns in human gait: Variability and compensating effects, *Hum Mov Sci*, **3**(1–2):51–76, 1984.
5. Lafortune MA, Cavanagh PR, Sommer HJ, Kalenak A, Three-dimensional kinematics of the human knee during walking, *J Biomech* **25**:347–357, 1992.

6. Li J, Wyss UP, Costigan PA, Deluzio KJ, An integrated procedure to assess knee-joint kinematics and kinetics during gait using an optoelectric system and standardized X-rays, *J Biomed Eng* **15**(5):392−400, 1993.
7. Chao EY, Laughman RK, Schneider E, Stauffer RN, Normative data of knee joint motion and ground reaction forces in adult level walking, *J Biomech* **16**(3):219−233, 1983.
8. Benedetti MG, Catani F, Leardini A, Pignotti E, Giannini S, Data management in gait analysis for clinical applications, *Clin Biomech* **13**(3):204−215, 1998.
9. Shiavi R, Limbird T, Frazer M, Stivers K, Strauss A, Abramovitz J, Helical motion analysis of the knee. II. Kinematics of uninjured and injured knees during walking and pivoting, *J Biomech* **20**(7):653−665, 1987.
10. Frigo C, Rabuffetti M, Kerrigan DC, Deming LC, Pedotti A, Functionally oriented and clinically feasible quantitative gait analysis method, *Med Biol Eng Comput* **36**(2): 179−185, 1998.
11. Isacson J, Gransberg L, Knutsson E, Three-dimensional electrogoniometric gait recording, *J Biomech* **19**(8):627−635, 1986.
12. Kadaba MP, Ramakrishnan HK, Wootten ME, Gainey J, Gorton G, Cochran GVB, Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait, *J Orthop Res* **7**(6):849−860, 1989.
13. Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, Renstrom P, *In vivo* knee kinematics during gait reveals new rotation profiles and smaller translations, *Clin Orthop Relat Res* **454**:81−88, 2007.
14. Watelain E, Barbier F, Allard P, Thévenon A, Angué JC, Gait pattern classification of elderly healthy men on the basis of biomechanical data, *Arch Phys Med Rehabil* **5**(81): 579−586, 2000.
15. Vardaxis VG, Allard P, Lachance R, Duhaime M, Classification of able-bodied gait using 3-D muscle powers, *Hum Mov Sci* **17**(1):121−136, 1998.
16. Toro B, Nester CJ, Farren PC, Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy, *Gait Posture* **25**(2):157−165, 2007.
17. Xu G, Zhang Y, Begg R, Mining gait pattern for clinical locomotion diagnosis based on clustering techniques, *Proc 2nd Int Conf Adv Data Mining Appl* 296−307, 2006.
18. Mulroy S, Gronley J, Weiss W, Newsam C, Perry J, Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following strok, *Gait Posture* **18**:114−125, 2003.
19. O'Byrne JM, Jenkinson A, O'Brien TM, Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer, *J Child Neurol* **13**:101−108, 1998.
20. Kienast G, Bachmann D, Steinwender G, Zwick EB, Saraph V, Determination of gait patterns in children with cerebral palsy using cluster analysis, *Gait Posture* **10**:57, 1999.
21. O'Malley MJ, Abel MF, Damiano DL, Vaughan CL, Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters, *IEEE Trans Rehabil Eng* **5**(4):300−309, 1997.
22. Riley PO, Paolini G, Della Croce U, Paylo KW, Kerrigan DC, A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects, *Gait Posture* **26**(1):17−24, 2007.
23. Hagemeister N, Yahia L'H, Duval N, de Guise JA, *In vivo* reproducibility of a new non-invasive diagnostic tool for three-dimensional knee evaluation, *Knee* **6**(3):175−181, 1999.
24. Lustig S, Magnussen R, Cheze L, Neyret P, The KneeKG system: A review of the literature, *Knee Surg Sports Traumatol Arthrosc*, **20**(4):633−638, 2011.
25. Bishop C, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

26. Stevens JP, *Applied Multivariate Statistics for the Social Sciences*, Routledge Academic, London, 1992.
27. Hubert M, Rousseeuw P, Verdonck T, Robust PCA for skewed data and its outlier map, *Comput Stat Data Anal* **53**(6):2264−2274, 2009.
28. Deluzio KJ, Wyss UP, Costigan PA, Sorbie CS, Principal component models of knee kinematics and kinetics: Normal *vs.* pathological gait patterns, *J Hum Mov Sci* **16**:201−217, 1997.
29. Muniz AMS, Nadal J, Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait, *Gait Posture* **29**(1):31−35, 2009.
30. Jolliffe IT, *Principal Component Analysis*, Springer, New York, 2003.
31. Edward Jackson J, *A User's Guide to Principal Components*, John Wiley & Sons, New York, 1991.
32. Hair JF, Anderson RE, Tatham RL, Black WC, *Multivariate Data Analysis with Readings*, Englewood Cliffs, Prentice-Hall, NJ, 1998.
33. Bellcore AB, Buja A, Eyuboglu N, Remarks on parallel analysis, *Multivariate Behav Res* **27**:509−540, 1992.
34. Robert H, *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*, CRC Press, New York, 2006.
35. Franklin SB, Gibson DJ, Robertson PA, Pohlmann J, Fralish JS, Parallel analysis: A method for determining significant principal components, *J Veg Sci* **6**:99−106, 1995.
36. Bandyopadhyay S, Mukhopadhyayand A, Maulik U, An improved algorithm for clustering gene expression data, *Bioinformatics* **23**(21):2859−2865, 2007.
37. Altman D, *Practical Statistics for Medical Research*, Chapman & Hall, New York, 1991.